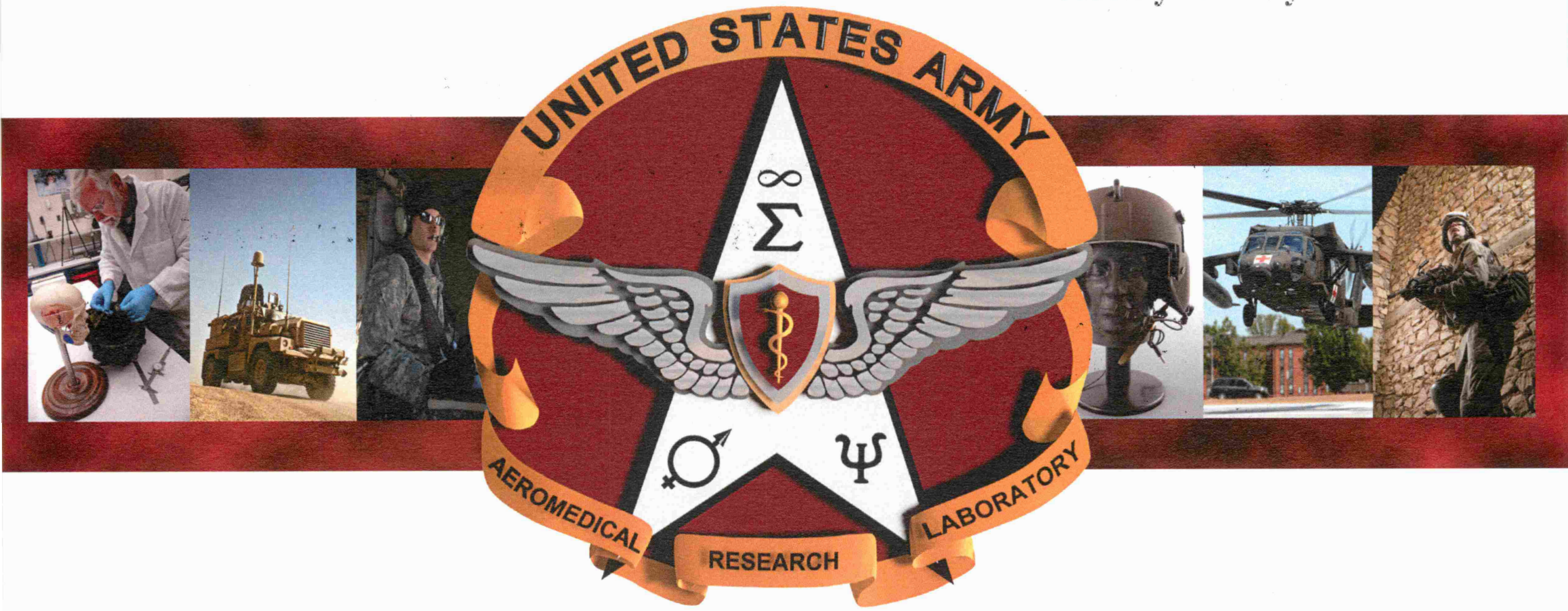
USAARL Report No. 2012-11

A Review of Computerized Team Performance Measures to Identify Military-Relevant, Low-to-Medium Fidelity Tests of Small Group Effectiveness during Shared Information Processing

By Ben D. Lawson
Amanda M. Kelley
Jeremy R. Athy



United States Army Aeromedical Research Laboratory

Warfighter Health Division

May 2012

Approved for public release, distribution unlimited.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
25-05-2012	Final	2010-2012

4. TITLE AND SUBTITLE	
A Review of Computerized Team Performance Measures to Identify Military-Relevant, Low-to-Medium Fidelity Tests of Small Group Effectiveness during Shared Information Processing	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
Ben D. Lawson Amanda M. Kelley Jeremy R. Athy	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
U.S. Army Aeromedical Research Laboratory P.O. Box 620577 Fort Rucker, AL 36362	USAARL 2012-11
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
U.S. Army Medical Research and Materiel Command 504 Scott Street Fort Detrick, MD 21702	USAMRMC
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release, distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Many tests have been developed to measure team performance, but no test has emerged as the clear leader and few attempts have been made to compare existing tests. This review identified the existing computerized measures of small group cognitive performance and determined which had the greatest potential for low-to-medium fidelity, military-relevant research. More than 50 potential measures were identified, among which the following were considered most suitable: Tactical Navy Decision-Making System (TANDEM), Team Performance Assessment Technology (TPAT), Team Interactive Decision Exercise for Teams Incorporating Distributed Expertise (TIDE2), C3 (Command, Control & Communications) Interactive Task for Identifying Emerging Situations (NeoCITIES), Agent Enabled Decision Group Environment (AEDGE), Distributed Dynamic Decision Making (DDD), and Duo Wondrous Original Method Basic Awareness/Airmanship Test (DuoWOMBAT). The advantages and disadvantages of each are described, which should aid researchers to identify the test most appropriate to their experimental needs.

15. SUBJECT TERMS

team performance, group performance, shared performance, performance measurement

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Loraine St. Onge, PhD
					19b. TELEPHONE NUMBER *(Include area code)*
UNCLAS	UNCLAS	UNCLAS	SAR	42	334-255-6906

Reset

Summary

This report identifies the available computer tests that are designed to assess the performance of multiple persons engaged in shared information processing tasks. The purpose of this report is to examine the literature concerning group cognitive performance measures and consider their potential for military research relevant to a small team, such as an infantry squad or the aircrew of an aircraft. Tests were favored which were quantitative and automated. Additionally, this review focused on tests which assessed real-time information processing (rather than physical performance or managerial skill), and which could be generalized across more than one mission, platform, or experiment. Five-hundred seventy-one potentially-relevant abstracts were reviewed; of which 73 full text articles were deemed relevant and 54 candidate measures of team performance were identified. After further review, seven tests were selected as the most appropriate to future military research within the stated scope of this report (as described in the Methods section). The seven tests of greatest interest were (listed in no order of preference): Tactical Navy Decision-Making System (TANDEM), Team Performance Assessment Technology (TPAT), Team Interactive Decision Exercise for Teams Incorporating Distributed Expertise ($TIDE^2$), C^3 (Command, Control, & Communications) Interactive Task for Identifying Emerging Situations (NeoCITIES), Distributed Dynamic Decision Making (DDD), Agent Enabled Decision Group Environment (AEDGE), and Duo Wondrous Original Method Basic Awareness/Airmanship Test (DuoWOMBAT). The characteristics, strengths, and potential limitations of each test are discussed briefly in this report and references are provided for further information. Each of these seven tests is designed to be relevant to the performance of military or paramilitary crews or teams. Only one test (NeoCITIES) is not designed specifically for military applications, but it was included because it has many desirable features and it is suitable for paramilitary (e.g., police) situations and for scenarios relevant to national defense, such as simulating a coordinated emergency response to terrorist attacks on civilian centers. The test most similar to the rudimentary aspects of flight control tasks engaged in by military aviation crewmembers is the DuoWOMBAT. The other six tests (TANDEM, TPAT, $TIDE^2$, NeoCITIES, DDD, and AEDGE) focused on various aspects of team performance most relevant to command/control situations, such as handling threats and allocating resources. The tests which were judged as most likely to be relevant, readily available, widely/recently used, and relatively mature in terms of validation included NeoCITIES, DDD, and DuoWOMBAT. Of these, AEDGE, DDD, and DuoWOMBAT are clearly available for immediate purchase. The Warfighter Health Division of the U.S. Army Aeromedical Research Laboratory (USAARL) currently owns a copy of DDD and DuoWOMBAT, which have been chosen to fill past or current research needs. A new test in development (C3Conflict) was identified after the completion of literature gathering for this review; it appears to have many desirable features and should be considered further as more validation work is done. Future research on military and paramilitary team performance should consider the information in this report when seeking to identify the tests most appropriate to the specific needs of the scientific effort being planned. Further use, refinement, validation, and comparison of the existing automated group performance measures are encouraged.

This page is intentionally left blank.

Acknowledgments

It is fitting that extensive teamwork is required to write a paper on team performance. We thank Jillian Parker, Melinda Vasbinder, Shauna Legan, John Golding, Micah Mobley, and Roger Cholewiak for their valuable assistance and comments during the preparation of the manuscript. We also thank Diana Hemphill, Sharon Fales, Jean Southwell, and Alexandria Austermann for their expert assistance in obtaining many of the abstracts and articles considered in this review. Last but not least, we thank Lori St. Onge, Arthur Estrada, Timothy Cho, Catherine Davis, and Dennis Bergstrazer for their detailed and attentive proofreading.

This page is intentionally left blank.

Table of contents

This page is intentionally left blank.

Introduction

Coordinated team performance is important to the success and safety of military personnel in nearly all of the missions they are asked to perform. Efficient shared processing of information is a critical feature of operations performed by units of infantry, armor, artillery, aviation, special forces, logistics, medical services, military intelligence, and communications personnel. The harmful effect of inaccurate shared cognition is well-documented as a contributing factor to aircraft mishaps (Salas et al., 2001; Katz et al., 2006) and command/control-related accidents (Armstrong, 1994).

Research on human performance traditionally focused on the individual, but has gradually expanded from a consideration of individual information processing to recognition of the role of shared cognition (Ntuen, 2006). As a result, computerized measures of team performance are being developed to quantify shared information processing. Good measures of team performance would benefit the military by allowing it to determine the characteristics of good teamwork and evaluate the effectiveness of team training methods (Baker and Salas, 1997). Additionally, good measures of team performance should aid the development of cost-effective training simulations (Simpson and Oser, 2003).

Although several computerized tests have been developed to measure military-relevant team performance, no test obviously dominates the field of inquiry in the way that measurement of individual performance has tended to be dominated by such tests as those found within the Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB) (Reeves and Thorne, 1986; Englund et al., 1985). There is limited information concerning which team performance tests are optimal for military small group research, and few reports have been published describing and comparing the available tests. The most recent comparative review was written five years ago by Go, Bos, and Lamoureux (2006), who thoroughly reviewed 44 potential test platforms. Prior to that, Banner (2004), reviewed seven team performance tests, while Bowers and Jentsch (2001) reviewed the suitability of 36 commercial computer games for use in team performance research.

The limited number and recency of comparative information on computerized team performance tests contributes to the present lack of uniformity in the measurement of military team performance and limits comparisons across studies. For these reasons, a literature review was performed to assess the current state of team performance measures and identify those most suitable for military research. This review focused on low-to-medium fidelity systems, since they tend to be more commonly-used and widely disseminated than high-end, custom-built, “one-of-a-kind” simulation facilities (Banner, 2004; Dahlstrom et al., 2009; Jentsch and Bowers, 1998; Bowers and Jentsch, 2001), due to their relative affordability, portability, configurability, and ease of adoption.

A number of approaches, including surveys, behavioral checklists, and computer tests or simulations, have been used to quantify team performance (Brannick, Salas, and Prince, 1997). This report focused solely on quantitative computerized tests of team performance on shared information-processing tasks. The purpose of this report was to identify a subset of the most suitable automated measures of team performance for use in military research. This report was not intended as a comprehensive review of the field of team research or the various issues surrounding measurement of team performance, but rather as a practical guide to the available computerized team performance metrics. The literature abounds with reviews of the history and state of team performance research, discussions of team coordination and team training approaches, discussions of problems of definition in team performance research, theoretical discussions of the essential nature and aspects of team performance, working models of team performance, discussions of the importance of measuring team performance, descriptions of the problems and issues to be considered when attempting to measure team performance, and recommendations concerning how to develop measures of team performance (e.g., what a good measure should be able to do). This report attempts to address a less-frequently discussed question, which can be phrased as follows: "What automated small team performance measures exist which I should consider applying to my military research studies?" This report seeks to answer that question by identifying the most appropriate tests for this purpose. It was not the purpose of this review to identify the "single best" test. Rather, the purpose of this review was to produce a short list of the most suitable tests for research on military team performance, in hopes that future research activity would become more focused and the list narrowed down further by the subsequent efforts of the research community. Greater standardization of team performance testing protocols would be of benefit to the military research community.

Methods

A literature search was conducted for articles published in 1990, or later. The searched databases included Defense Technical Information Center, PsycInfo, PubMed, and Army Research Laboratory (ARL) library online. The primary search terms were "team" and "performance." Additional search terms were discussed with professional librarians, who assisted with the literature search. The librarians based their search on an idealized set of criteria exemplified by the following compound statement devised by the first author:

[Objective / Performance-Based / Quantitative /Automated / Computer / Computerized…
(NEAR)
…Measure / Metric / Battery / Test / Assessment / Task / Technology]
(AND)
[(Team / Group / Shared / Squad / Crew / Cockpit) Performance…
(NEAR)
…Effectiveness / Cognition / Cognitive / Decision / Attention / Coordination / Resource Management]

A team of three research psychologists (the report authors) reviewed the resulting literature obtained by the search. The authors excluded those items that were not clearly related to the focus of this review. The qualitative exclusion criteria included the following:

a. Items were excluded which did not describe the measurement of team cognition or performance.

b. Items were excluded which dealt solely with business/management issues, such as team building, personnel selection, or leadership (i.e., we excluded reports where such variables were not part of a group of other relevant measures being assessed directly by the group's keyable responses during a computerized performance test).

c. Items were excluded concerning physical performance studies (e.g., strength, endurance) if they did not also include a cognitive or psychomotor aspect.

d. This review sought to identify computerized performance tests which could be adapted to current research efforts, so literature prior to 1990 was excluded, since articles before that time would have preceded the mainstream use of personal computers, graphical user interfaces for Microsoft © Windows, hyper text markup language, etc. However, in a few cases, some important post-1990 articles were obtained on team performance tests which referred back to initial development efforts of the same test shortly before 1990, in which case we provide those earlier references in this report.

A collection of abstracts or articles was identified for further evaluation. After evaluation of these items, full text articles were obtained of the most relevant items selected for a second round of more detailed evaluation. The first and second round of evaluations gave special preference to articles which came closest to meeting the following inclusion criteria:

a. Reports were preferred which described systematic, quantitative, computerized performance tasks (as distinguished from surveys, observer-based methods, task analyses, theoretical papers, computer games, or training improvements or guidelines).

b. Reports were preferred which described military-relevant tasks and measures (vice measures only applicable to non-military situations or general measures of cognitive state). Greatest consideration was given to tasks/measures which appeared to be part-task or medium-fidelity simulations similar to many tasks which must be performed by small military groups (e.g., squads, aircrew).

c. Reports were preferred which were relevant to small groups (vice entire agencies or companies); with a small group being defined as 10 people or fewer (4 to 10 people are considered a "squad" by the military).

d. Reports were preferred which were relevant to real-time shared information processing (as distinguished from long-range or strategic planning). For example, the first author would include any report concerning the speed and accuracy of a small group's ability to detect target events while communicating and coordinating an appropriate response.

e. Reports were preferred which described generalizable tasks or tests not limited to one mission, platform, experiment, project, facility, or course of training/simulation. For example, we excluded unique simulations of high complexity and cost (such as the 150-person Virtual Warfare Center) which are likely to exist in only one or two places. Similarly, training-only applications or games with limited scoring outputs that are not intended primarily for scientific research were excluded (e.g., the Army's laptop-based OneSAF ground warfare training game).

f. We did not seek to identify every generation of a given test throughout its history of development. Rather, we included only the more recent versions. For example, NeoCITIES derived from CITIES, but we only list NeoCITIES in this review. Similarly, TANDEM is derived from the earlier Team Performance Assessment Battery (TPAB), and both the TPAB and the Team Resource Management Test (REMAN) are derived from still earlier tests such as the Multiple Task Performance Battery (MTPB) and Distributed Resource Allocation and Management (DREAM) task (Bowers, Urban, and Morgan, 1992). For the purposes of this review, we have listed TANDEM, TPAB, and REMAN below, but not the earliest tests such as MTPB or DREAM. It should be noted that this type of exclusion was rare, since most tests have kept their original name as they were modified or have only changed names once.

Results and Discussion

Round 1 findings: Identification of an initial list of potential tests

During the initial literature search, 571 abstracts or articles were identified for preliminary review, from which 73 full text articles were deemed to merit a second round of more detailed review. Among these 73 reports, 54 potential team performance tests were identified. This full list of potential measures is shown below. The reader should note that many of the items below did not turn out to be computerized tests of team performance once the exclusion and inclusion criteria were more carefully assessed during further review (see Round 2 findings below). Below are the names of each item and one or more resource citations.

a. Team Performance Assessment Battery (TPAB)
   (1) Bowers, Urban, and Morgan (1992)
   (2) Urban et al. (1995)
   (3) Schraagen and Rasker (2003)

b. Team Performance Assessment Simulation (TPAS)
   (1) Swezey et al. (1998)

c. Team Performance Assessment Technology (TPAT)
   (1) Swezey, Hutcheson, and Swezey (2000)
   (2) Lamoureux et al. (2006) (see appendix p A-13)

d. Tactically Relevant Assessment of Combat Teams (TRACTs)
   (1) Fowlkes et al. (1999)

e. Team Resource Management Test (REMAN)
   (1) Woldring and Issac (1999)
   (2) Hintz (2011)

f. Tactical Navy Decision-Making System (TANDEM)
   (1) Canty and Schwab (2001)

(2) Dwyer et al. (1992)
(3) Van Berlo (2004)
(4) Lenox et al. (1999)

g. Decision-making Evaluation Facility for Tactical Teams (DEFTT)
   (1) Johnston, Poirier, and Smith-Jentsch (1998)

h. Tactical Decision-Making Under Stress (TADMUS)
   (1) Cannon-Bowers and Salas (1998)
   (2) Johnston, Cannon-Bowers, and Salas (1998)

i. Adaptive Architectures for Command and Control (A2C2)
   (1) Entin and Entin (2000)

j. Tactically Relevant Assessment of Combat Events (TRACE)
   (1) McCluskey et al. (1998)

k. Team Interactive Decision Exercise for Teams Incorporating Distributed Expertise ($TIDE^2$)
   (1) Hollenbeck et al. (1991)
   (2) Hollenbeck et al. (1997)
   (3) Hollenbeck et al. (1995)
   (4) Lamoureux et al. (2006) (see appendix p. A-17)

l. Targeted Acceptable Responses to Generated Events or Tasks (TARGETs)
   (1) Fowlkes et al. (1992)
   (2) Fowlkes et al. (1994)

m. The Army Command and Control Evaluation System (ACCES)
   (1) Hayes, Layton, and Ross (1993)
   (2) Halpin (1996)
   (3) Essens et al. (2005)

n. NeoCITIES (derived from CITIES)
   (1) Jones et al. (2004)
   (2) McNeese et al. (2005)

o. Hierarchical Task Analysis (Teams) (HTA[T])
   (1) Annett, Cunningham, and Mathias-Jones (2000)

p. Duo Wondrous Original Method Basic Awareness/Airmanship Test (DuoWOMBAT)
   (1) Breton, Tremblay, and Banbury (2007)

q. Tactical Simulation System (TACSIM)
   (1) Tactical simulation system (TACSIM) (2011)
   (2) Go, Bos, and Lamoureux (2006) (see appendix p. A-10)

r. Distributed Dynamic Decision Making (DDD)
   (1) Galster, Nelson, and Bolia (2005)
   (2) Aptima, Inc. (2005)

s. DDD-III Simulator: a) North African “Insertion from the Sea;” b) Joint Command/Control; c) Joint Task Force; d) Joint Task Force Group

(1) Wollenbecker et al. (1999)
(2) Hocevar et al. (1999)
(3) Levchuk et al. (1999)
(4) Lamoureux et al. (2006) (see appendix p. A-21; A-30; A-33; A-35)

t. DDDnet, Airborne Warning and Control System (AWACS) Weapon Director Teams

(1) Barnes, Elliott, and Entin (2001)
(2) Lamoureux et al. (2006) (see appendix p. A-31)

u. Agent Enabled Decision Group Environment (AEDGE)[1]

(1) Barnes et al. (2004)
(2) Barnes et al. (2004)
(3) Elliot et al. (2002)
(4) Barnes, Petrov, and Elliott (2002)

v. Cognitive Engineering Research on Team Tasks (CERTT)

(1) Cooke and Shope (2005)
(2) Cooke (2002)
(3) Go, Bos, and Lamoureux (2006), (see appendix p. A-15)

w. TEAMSim

(1) DeShon et al. (2004)
(2) DeShon, Brown, and Greenis (1996)

x. Air Combat Mission Planning (ACMP)

(1) Gaylord and Sowell (1992)

y. Anti-Air Teamwork Observation Measure (ATOM)

(1) Smith-Jentsch et al. (1998)
(2) Shanahan et al. (2007)
(3) Entin and Entin (2001)

z. Low-Fidelity Aviation Research Methodology (LFARM)

(1) Bowers et al. (1992)

aa. Controller Teamwork Evaluation and Assessment Methodology (CTEAM)

(1) Bailey et al. (1999)

bb. Roboflag

(1) Funke and Galster (2009)
(2) Guznov et al. (2011)

[1] This test is referred to by various phrases and acronyms, with “AEDGE” and “GROUP” being most common.

cc. Space Fortress
  (1) Shebilske et al. (1999)

dd. Advanced Disaster Management Simulator (ADMS)
  (1) This system is described at: http://trainingfordisastermanagement.com/
  (2) Go, Bos, and Lamoureux (2006) (see appendix p. A12).

ee. Air Operations Centres (AOC), AWACS in the Command, Control, and Communications Simulation, Training and Research System (C3STARS) Facility
  (1) This system is described at:
  a) ww.hec.afrl.af.mil/Organization/HECP/AOC.asp
  b) www.mesa.afmc.af.mil/html/c3stars.htm
  (2) Go, Bos, and Lamoureux (2006) (see appendix p. A-4; A-5)

ff. NASA Ames Center – Distributed Facilities
  (1) Jonas (2008)
  (2) National Space Biomedical Research Institute (2010)
  (3) Go, Bos, and Lamoureux (2006) (see appendix p. A-7)

gg. One Semi-Automated Forces (OneSAF)
  (1) OneSAF Objective System is described at: www.onesaf.org/onesaf.html
  (2) Go, Bos, and Lamoureux (2006) (see appendix p. A-9)

hh. Virtual Warfare Centre (VWC)
  (1) Villaneuva (2007)

ii. Synthetic Task Environment (STE) in Cognitive Engineering Research on Team Tasks (CERTT) Lab
  (1) Lamoureux et al. (2006) (see appendix p. A-4)

jj. Team and Individual Tactical Assessment Network (TITAN)
  (1) Blais, Thompson, and Baranski (2002)
  (2) System described at: http://ntt.ca/

kk. Bolo, Tank Battle Game
  (1) Knight, Durham, and Locke (2001)

ll. Dangerous Waters, Naval Combat Experience
  (1) Commercial site for Sonalysts Combat Simulations: www.scs-dangerouswaters.com
  (2) Go, Bos, and Lamoureux (2006) (see appendix p. A-29)

mm. Longbow 2, Helicopter Flight Simulator
  (1) Marks et al. (2002)

nn. Team Argus, a radar-like classification task
  (1) Miller (2001)
  (2) Schoelles and Gray (1998)

(3) Lamoureux et al. (2006) (see appendix p A-19)

oo. Neverwinter Nights: a) Recovering weapons from hidden caches; b) Capture the Flag
   (1) The commercial website is: http://www.bioware.com/games/legacy
   (2) Weil et al. (2005)

pp. ATC team training device
   (1) Bailey and Thompson (2000)

qq. Multi-agent Operation Range Simulation Environment (MORSE)
   (1) Rectenwald et al. (2003)
   (2) Sycara et al. (2005)

rr. SCUDHunt
   (1) Perla et al. (2000)
   (2) Holzworth (2002)

ss. Wright State Aegis Simulation Platform (WASP)/Team Aegis Simulation Platform (TASP – an extension of WASP)
   (1) These systems are described at: http://www2.ie.psu.edu/Rothrock/Research/HPAM/
   (2) Lamoureux et al. (2006) (see appendix p. A-26)

tt. Janus Wargaming Simulation
   (1) Chapman et al. (2002).

uu. Networked Fire Chief (NFC)
   (1) Chapman et al. (2002)
   (2) Chapman (2000)

vv. C3Fire
   (1) Dube et al. (2010)
   (2) Granlund (2003)
   (3) Persson and Worm (2002)

ww. Microsoft Flight Simulator, Pilots (ATC)
   (1) Brannick et al. (1995)
   (2) Lamoureux et al. (2006) (see appendix p. A-34)

xx. Ruthless.com (Red Storm Entertainment)
   (1) Bowers and Jentsch (2001)

yy. Fleet Command (Jane's Combat Simulations). Note: Of four games listed by Bowers and Jentsch (2001), this one simulates a fairly modern and realistic combat scenario (involving naval tactics).

(1) Bowers and Jentsch (2001)

zz. Half-Life Team Fortress Classic (Valve)
   (1) Bowers and Jentsch (2001)

aaa. Antietam (Fireaxis Games).
   (1) Bowers and Jentsch (2001)

bbb. Beer Distribution Game
   (1) Banbury et al. (2010)
   (2) Goodwin and Franklin (1994)
   (3) Geyer-Schulz (1996)

The initial list of candidate items above were evaluated via the exclusion/inclusion criteria and narrowed down to a list of the seven most relevant tests, which are described below.

Round 2 findings: Identification and description of the most suitable tests

Several trends were noticed during Round 2 of the literature review. First, the majority of the initial literature matching the stated search terms (see Methods) led to business or management items, many of which were opinion pieces of a philosophical, inspirational, or otherwise non-scientific nature. Second, many of the initial hits which appeared to discuss tests (and therefore to be pertinent to this review) did not actually yield full-text articles describing generalizable tests or test batteries, but rather, descriptions of research projects, laboratory facilities, or the general problems surrounding the measurement of team cognition. Third, when a report described a potentially relevant team performance test (or tests), practical information was sometimes insufficient to make inferences concerning the ease of test administration, time required for testing, ease of access (e.g., is it open-access or available commercially "off-the-shelf?"), maturity of the test (e.g., is it widely used, established, reliable, and valid?), extent of automated and objective scoring, generalizability, and configurability (to different tasks or team sizes). Additionally, for some of the older reports, it was difficult to determine (by additional web searches or e-mail inquiries) whether the test was compatible with the latest hardware/software, whether there was continued development and use of the test, whether the report described the latest version of the test, or whether the test is available for use or purchase. The lack of practical information on performance tests has been a problem for human performance measurement in general and has made applied sources such as the Human Performance Measures Handbook (Gawron, 2000) particularly useful.

Despite these challenges, the full list of more than 50 potential measures was considered against the aforementioned qualitative inclusion/exclusion criteria and the following tests were selected unanimously as the most interesting for continued evaluation in military team performance research (tests are listed in random order):

a. TANDEM

b. TPAT

c. TIDE[2]

d. NeoCITIES

e. DDD

f. AEDGE

g. DuoWOMBAT

To corroborate this review, preliminary findings were presented at two professional conferences (Estrada et al. 2010; Lawson, Kelley, and Athy, 2011) attended by various human performance researchers, including people knowledgeable about team performance. The audiences were invited to recommend any additional tests of importance which may have been overlooked by this review effort. No additional tests were suggested to the authors at these two conferences.

Some trends were noticed concerning the seven tests which were selected as most suitable for military research. First, they tended to focus on the general performance of the team, rather than on specific and established aspects of neurological or cognitive functioning. Second, almost all of the tests involved simulations of command/control tasks.

Some authors or laboratories contributed disproportionately to the number of potential tests on our initial list of more than 50 items. In such cases, only the author's most recent, active, or mature tests were selected for inclusion in the final list of seven. For example, researchers at the Institute for Simulation and Training at the University of Central Florida (and their colleagues) have been very active in teamwork research, and have come up with a number of team performance research projects, laboratories, simulations, testbeds, team tasks, or team tests, including DREAM, LFARM, TARGETS, TPAB, REMAN, and TANDEM. Of these, the authors selected TANDEM for inclusion in the final list of seven most suitable computerized team performance tests.

The authors reviewed the tests for trends concerning the most common independent variables, and found that workload was the most common factor, which could be manipulated by the experimenter during administration of the tests (e.g., via manipulation of the number of stimuli to keep track of and/or the duration of time to respond).

Below, we briefly summarize each of the seven most interesting tests, providing one or more suitable references for further reading, describing the main tasks performed by the subjects, providing an example of an operational scenario relevant to the test, listing the main variables being studied by the test (i.e., measured or manipulated), and describing the potential strengths and limitations of the test. Since the amount and type of information available on each test was not the same, the summary information about each test will vary in the sections below.

Tactical Naval Decision-Making System (TANDEM)

Developed by the Naval Training Systems Center (known since 1993 as the Naval Air Warfare Training Systems Division), TANDEM is a low-fidelity simulation of a command, control, and communications environment (Dwyer et al., 1992). TANDEM was created to simulate a military combat information center. It requires subjects to identify various targets and decide upon an appropriate response. Following target identification, the subjects are to perform a "final engagement" which involves a "shoot" or "clear" decision (Dwyer et al.). The identification of the target is made by various indicators that may clearly define the target's characteristics or may be ambiguous/contradictory, requiring the crew to engage in further interpretation and coordination.

When the simulation begins, several unidentified targets are presented on a screen similar to that found on a radar operator's display. The team must select (or "hook") a target of interest, which will be presented with potentially important information concerning the target. Typically, two or three team members work together and each member is given different information about the target. For example, one team member may get information on the target's type (aircraft, ship, or submarine), another on the target's classification (military or civilian), and a third on the target's likely intent (peaceful or hostile) (Lenox et al., 1999). From this information, either a team vote or a team leader will determine what the final engagement decision will be (depending on the scenario). Following engagement, the target is typically removed from the screen. Performance measures include such variables as the "hook time" (how long a target is selected), accuracy of labeling of the target, and the number of targets engaged, but variations of the task may include other measures such as whether a target reached a dangerous proximity region (represented by a small inner circle within the display).

System requirements for TANDEM are very basic. It is personal computer-based and programmed in C++. TANDEM requires at least a 640 by 480 video graphics array (VGA) based graphics card, a 20 megabyte hard drive, 640 kilobytes of random-access memory (RAM), a Logitech three-button trackball (also works with a mouse) with 4.01 driver (or equivalent), and Microsoft DOS version 3.3 or later. The TANDEM task measures the participant's memory, decision making ability, and the interdependence of team members, and can vary the task by influencing variables such as the overall workload and the ambiguity of the information presented. TANDEM is fully automated, configurable, and is militarily relevant. TANDEM lacks a team-of-teams capability, and according to Weaver et al. (1995), its largest potential shortcoming is a failure to require the integration of dynamic information over time. Nevertheless, this limitation could make the test attractive for those researchers wishing to do a simple study wherein stimuli and communications are more limited and controlled. Several TANDEM study authors were contacted, but no information could be obtained concerning purchasing or otherwise obtaining the TANDEM software. As a government-sponsored product developed for a Department of Defense (DoD) laboratory, it may be available without cost for legitimate research uses by government personnel.

Summary of TANDEM

a. Recommended reading: Dwyer et al. (1992); Canty and Schwab (2001); and Cannon-Bowers and Salas (1998).

b. Team task: Command/control; two to three participants; members determine type, threat level, and intent of each contact.

c. Scenario: Military combat information center.

d. Main factors studied: Decision-making, including interdependence, time pressure, workload, and ambiguity.

e. Strengths: Automated; configurable; generalizable; militarily-relevant.

f. Potential limitations: Relatively low fidelity (does not require integration of changing information over time); limited on specific aspects of cognitive ability or shared knowledge; no "team-of-teams" capability; availability uncertain.

Team Performance Assessment Technology (TPAT)

Developed by InterScience America, TPAT assesses individual and team performance in a group task. TPAT measures some higher-order activities, unlike many other team tasks that are based more on the learning of patterns or rules. Training is provided by tutorials that require the user to reach a criterion of understanding. Once all team members reach their baseline, the main task may begin.

The scenario-based task involves a team of individuals that must deal with constant changing of information over time, for example, cooperating to extract a hostage from a fictitious hostile nation or cooperate as nuclear power plant technicians. Each person has control over a certain aspect of the task, and as events unfold, the team is required to make decisions on what steps should be taken next. Besides the events that always occur, TPAT also uses the decisions made by the teams to create new situations, thus requiring new decisions to be made. This allows for a quasi-experimental design. Overall, over 1000 decision alternatives exist (Swezey, Hutcheson, and Swezey, 2000).

Teams consist of up to nine team members, with the nine members generally split into three separate units of three individuals. A command structure exists within the units and for the entire group of units. If nine team members are not available, the TPAT program is able to replicate any missing team members, allowing testing of one to nine individuals in this task. Information is provided to team members via computer or team member messages concerning events occurring within the individual's domain of labor. Based on these messages, team members are required to make preliminary decisions on what should be done. While making these decisions, team members are to indicate what earlier events influenced the decisions they are making and what possible future plans they may have. TPAT records the decisions acted upon and how they connect to previous events, letting researchers witness what led to a team's decision, and what future plans were not acted upon.

Scores are given for the team's performance and for the performance of each team member. No knowledge concerning computer programming is required to score TPAT, as the output software of the program provides a detailed evaluation of the participants. A total of 53 scores are provided, which include several performance factors (e.g., decision-making, planning, strategy development) and social psychological traits (e.g., communication, cohesion). System requirements for TPAT are very basic, and require networked Microsoft© Windows 95 or 98 compatible computers. Unfortunately, the first author of TPAT (Robert Swezey) died in 2002 (Van Cott, 2004) and little information is available about the company, InterScience America. At this time, TPAT may not be available or used extensively in research.

Summary of TPAT

a. Recommended reading: Swezey, Hutcheson, and Swezey (2000).

b. Team task: Three teams of three each; controlling command; air resources; and ground resources.

c. Scenario: One example is search/rescue and hostage extraction.

d. Factors studied: Decision-making; planning; strategy; situation awareness; initiative; communication; cohesion; leadership; task difficulty; task performance.

e. Strengths: Assesses group and individual performance; includes > 50 measures; accommodates up to nine users; has team-of-teams capability.

f. Potential limitations: First author deceased and little information available about this company. Uncertain if this test is still being actively developed/supported or sold commercially. The last available contact information for InterScience America (c. 1998) was 703-779-8090, Sterling, VA (also listed as being located in Leesburg, VA). When the first author called this telephone number, he reached what sounded like a home phone, as inferred from the answering machine message left by a small child.

Team Interactive Decision Exercise for Teams Incorporating Distributed Expertise (TIDE^2)

Developed by Hollenbeck et al. (1991), the TIDE^2 is a low-fidelity simulation of requiring classification of targets in ambiguous situations. The task relies heavily on Brunswik's lens model (Brunswick, 1940, 1943, 1955, and 1956) for individual decision making, but instead of providing all information to one individual, TIDE^2 requires a team of four individuals to make the classification. In order to accomplish this, three of the team members are "experts" within a given identification, while the fourth member is the team leader and has the final decision on target classification. This team organization of experts and a decision maker is found in many fields, and is the reason why TIDE^2 has been used in scenarios that resemble Naval command/control team tasks, personnel selection team tasks, and medical decision making team tasks (to name a few applications). The key to the TIDE^2 team task is that no individual has an understanding of all of the characteristics required to make all classifications of the targets, but among all of the team members, an accurate decision can be reached.

Although many different scenarios can be used for TIDE^2, we will discuss the Naval command/control type task, since that is most relevant to the military community. Each target typically has six characteristics that determine the threat level of the target. Despite the fact that all of the characteristics of the target are presented to all four team members, because of differences in pre-task training only "experts" can process certain characteristics of the target (usually one third of the information, thus two characteristics) and therefore all experts' opinions are needed to make an accurate decision. All team members are stationed at physically isolated computers connected by a network. This prevents verbal communication and allows for short messages only to be sent from one team member to another, allowing easy recording of communications by the researchers. The task begins with a "target" (such as an unknown aircraft) entering the team's airspace. One target is presented at a time in this task. A timer indicates how long the team has to designate the target's threat level. Each team member determines the target's threat value based on information within his/her domain of expertise, and then relays this to other team members (typically the leader) as he/she deems fit. Team members may also request information on the target. Targets can either contain information that clearly agrees across all characteristics (all point to the target being threatening), can be ambiguous (data falls in a grey zone for the experts), or can have data that is conflicting (expert one determines the target is a threat while experts two and three process the target as being non-threatening). Before the timer reaches zero, the "commander" must rate the target's threat level on a seven-point scale (ranging from the decision that the target can be safely ignored to the decision that a defensive action is necessary). Scoring of team performance entails giving two points for accurately diagnosing the target, one point for being one classification away from the correct target classification (selecting a two on the threat level when the target was a three), zero points for being two classifications away, minus one point for being three classifications away, and minus two points for being four or more classifications away. Feedback is given immediately after the final decision is made. If a decision is not made in the time allotted, the target is classified as missed or ignored.

For TIDE^2, experimenters can manipulate the time that a team has available to make the classification of the target, the time given in between the presentations of targets, and the agreement between the team members in the task. This allows researchers to create high stress scenarios (little time to make a decision, rapidly having to make decisions, or experts sending conflicting information to the leader) that can lead to limited communications or undervaluing certain team member's input. However, this task does not present changing information since all information is constant with regards to a given target.

The TIDE^2 teamwork task requires four very low-end IBM compatible computers with at least a 386 processor, DOS 4.01 or higher capabilities, and basic network connections to each other (the main computer should be the server). It should still be possible to purchase TIDE^2 software by contacting the original authors (Hollenbeck et al., 1991), who are associated with Michigan State University. The software was not expensive in 1991, selling for $25.00 at that time. The authors state that the software may be free for certain organizations. As a government-sponsored product developed originally for the Office of Naval Research (Hollenbeck et al.), it may be available without cost for legitimate research uses by government personnel. It is not certain if the software has been maintained and updated. Inquiries can be made to Dr. John Hollenbeck at jrh@msu.edu.

Summary of TIDE^2

a. Recommended reading: Hollenbeck et al. (1991); Hollenbeck et al. (1997).

b. Team task: Four team members (leader, etc.) must discover intent of targets.

c. Scenario: Naval command/control.

d. Factors studied include: Judgment accuracy; policy capturing (determining which aspects of the information are driving the ultimate decision); process tracing (the information-seeking process of participants).

e. Strengths: Military relevant; some validation work has been done.

f. Potential limitations: Uncertain if software is available commercially or has been updated.

C^3 (Command, Control, and Communications) Interactive Task for Identifying Emerging Situations (NeoCITIES)

The NeoCITIES task was developed to measure team performance, communication, and team cognition under pressure. The inception of the simulation/game was first described by Wellens and Ergener (1988). This primary article described a simulation task that allowed the experimenter to manipulate a multitude of independent variables while simultaneously and automatically recording a large number of dependent measures. Over time, the task evolved with improvements in technology and empirical knowledge. At present, the task employs a virtual city in a crisis management scenario requiring response from emergency services. The team members' shared goal is to respond appropriately to emergency events and prevent city-wide devastation while maintaining civil order. A total of two to three teams of two people each must communicate and work cooperatively to achieve these shared goals. Each pair is composed of an information manager and a resource manager. Communication options are varied, occurring by means of interactive touch-screens, monitors, microphones, and audio conferencing. Measurement of communication is possible between teams and also within teams. Each team must monitor changing events as well as the allocation of resources. The quantitative outcome variables in the task include communication frequency and type. Data are recorded electronically and can be supplemented and enriched by means of a full system structure including heart monitors on the participants.

The same researchers at Pennsylvania State University who developed NeoCITIES in its present form are moving this task further into the realm of modeling of dynamic decision making by advancing the structural components of the scenario employed. For example, these researchers are developing an interactive scenario for training and experimental assessment in NeoCITIES (e.g., El-Nasr, Jones, and McNeese, 2004).

NeoCITIES is a work in progress. The authors of this review were unable to locate specific information concerning test properties such as reliability and validity. However, it should be noted that the face validity of this measure appears adequate. The test is not specific to military applications, but it has relevance to some military operations. While some of the measures in this task are automated, the non-automated nature of the audio/video scoring method requires additional equipment and makes rapid, accurate analysis of communication data more challenging.

Currently, NeoCITIES is not available commercially, because it is a tool used in the academic setting and the developers cannot ensure maintenance and updates to multiple outside users. However, the test has been used in a number of settings and interested readers should contact Dr.

Michael McNeese (MMcNeese@ist.psu.edu) of Pennsylvania State University to determine whether research collaborations are possible. Presently, the program is being used by the Multidisciplinary Initiatives in Naturalistic Decision Systems (MINDS) Group at Pennsylvania State University (http://minds.ist.psu.edu), and by the North-East Visualization and Analytics Center, which is involved in an updated version called NeoCITIES Geo-Tools (www.geovista.psu.edu/resources/flyers/NEVAC_Thrust-4a_NeoCITIES_final.pdf).

Summary of NeoCITIES

a. Recommended reading: Wellens and Ergener (1988); El-Nasr, Jones, and McNeese (2004); project was last headed by McNeese (McNeese et al., 2005).

b. Team task: Monitoring of changing events; allocation of resources; two to three teams (fire, police, etc.) of two members each (information manager, resource manager).

c. Scenario: Medium fidelity metropolitan crisis control center; goal is to respond to emergency events, maintain order, and prevent city-wide catastrophe (e.g., due to terrorist attack).

d. Factors studied: Distributed decision-making; inference accuracy as a function of crisis tempo, data rate, and decision complexity.

e. Strengths: Allows study of teams-of-teams; has good measurement capabilities; was actively used and updated as of 2008; was one of the top-rated platforms (among 44) reviewed by Go, Bos, and Lamoureux (2006).

f. Potential limitations: Not a military scenario (but fairly applicable); limited information on shared knowledge; not all measures automated appear to be automated (e.g., audio/video); not available commercially.

<u>Distributed Dynamic Decision Making (DDD)</u>

The DDD task was developed by Aptima, Inc. to study aspects of team performance including communication in a complex and dynamic environment through a simulation platform. Validation efforts have been underway for 20 years and the task has been employed by military researchers for over 10 years (Aptima, 2010). DDD is marketed as a research tool that can also be used for training. While originally designed as a simulation of a military command/control environment, the task can be tailored to other contexts. Workload, information availability, and team structure can be manipulated. The dependent variables measured by the DDD represent individual and team performance and include, but are not limited to, “latency to process a task, accuracy in processing a task, percentage of tasks processed, and percentage of tasks processed at 100% accuracy” (Baker et al., 2004).

More than a dozen publications are listed on the Aptima website, which references and supports the DDD task. A trial version of the program is available for free download, as are Adobe© Portable Document Format (PDF) versions of tutorials and details of the specifications and requirements to use the DDD task. The software program operates using Microsoft© Windows XP, and data is logged in Extensible Markup Language (XML) format which simplifies the process of exporting to third party programs (e.g., for statistical analysis). Users have the freedom to customize their own scenarios and training environments. The program appears user-friendly and provides multiple resources to demonstrate and guide installation and configuration. However, the testing environment is minimally realistic, which may be an important consideration for certain types of research.

DDD has been adapted to an Internet-based version which allows participants in different locations to participate in the same mission in real-time. Using this feature, a maximum of 50 participants can engage in the same mission and "chat" using private or broadcast chat groups, e-mail communication, and voice-communication. It should be noted that for purposes of analysis, voice-communications are then scored by the researchers rather than automatically. The DDD task is designed to be compatible with other Aptima behavior modeling software for researchers who aim to further enrich the performance measurement capabilities of the program. Detailed review, analysis, and discussion of the DDD task are provided in MacMillan et al., (2004). USAARL recently purchased a configuration of DDD which cost under $20,000 (government price), but understands that further discounts may apply for academic agencies. For a DDD brochure, trial version, or price quoting information, see http://www.aptima.com/products/ddd.

Summary of DDD

a. Recommended reading: Galster, Nelson, and Bolia (2005).

b. Strengths: Measures automated; can easily manipulate task load/demand upon users (e.g., number, type, timing, uncertainty of tasks); can manipulate authority levels, communication, and information availability; much research activity over more than a decade.

c. Potential limitations: Small team size; tests fewer aspects of cognitive ability than some other tests (e.g., AEDGE); no "team-of-teams" capability.

Agent Enabled Decision Group Environment (AEDGE)

The AEDGE measure of team decision making employs a command/control scenario involving the weapons director team of an AWACS. This scenario embodies the core characteristics of a command/control environment including surveillance and communication. In this task, participants must exchange, interpret and weigh information as well as coordinate tactical action to successfully accomplish overall goals of the task (Go, Bos, and Lamoureux, 2006).

AEDGE is a Java-based technology which was developed by software engineers and researchers at 21st Century Systems, Inc. for training and performance research (Barnes et al., 2002). The development of the task platform involved the input of subject matter experts (SMEs), focus group interviews, and cognitive task analyses to ensure that the product was representative of operational settings. The task involves human users and computer-generated agents that may adopt any role in a scenario. Any entity (either friendly or hostile) not controlled by a human is controlled by the computer. A computer agent will make recommendations for a course of action which the human may or may not choose to view. Regardless of the human's performance and behavior, the system logs and captures the agent's recommendations thus allowing direct comparison between human and agent with respect to decision making and judgments. This enhances the system's ability to model individual and team performance. An additional benefit to the utilization of the agent-generated-action-recommendations is that the system may manipulate the quantity and quality (complexity) so as to vary the degree of cognitive workload. The experimenter has control over the configuration of a decision aide agent with respect to decision making style including, but not limited to, the degree of risk acceptable, aggressiveness, or certainty. Likewise, the experimenter may control the probability of a successful decision to be made given the environment. Dependent measures include communications expressed by speech as it is captured using voice generation technology and recordings. Other dependent measures are not as clearly stated in the references and resources available. However, the authors infer that the latency and accuracy of the decisions made are recorded.

This system seems to be well-suited for the researcher interested in how one person's decisions are influenced by the decision making style of a "partner" in a team scenario. Many of the options that may be manipulated by the experimenter are specific to the type of recommendations made by the agent-controlled entity (e.g., degree of directional bias, degree of riskiness, and degree of certainty). The product is available as commercial off-the-shelf in two variations and 21st Century Systems, Inc., offers maintenance and support services. Inquiries can be made to awilson@21csi.com.

Summary of AEDGE

a. Recommended reading: Barnes et al. (2002); AEDGE website, http://www.21csi.com/aedge.

b. Team task: High fidelity strategic/operational task – command/control.

c. Scenario: Weapons director roles in an AWACS system.

d. Factors studied: Individual and team workload; communication; decision-making.

e. Strengths: Has voice recognition and response; can monitor and vary communication frequency and media; can examine individual performance in the team setting; accommodates medium-size teams; AWACS simulation task based on SMEs and task analyses; appears readily available and actively used.

f. Potential limitations: Limited information on shared knowledge; possible mix of computer and observer measures; possible complexity.

<u>Duo Wondrous Original Method Basic Awareness/Airmanship Test (DuoWOMBAT)</u>

The memorably-named DuoWOMBAT is a modified version of the single-user WOMBAT and is designed to extend the WOMBAT to measure crew coordination and shared situation awareness. Participants must work cooperatively to accomplish relatively simple shared tasks. The test measures performance with respect to divided attention among multiple information sources, judgment of priorities, ability to estimate probable outcomes, judgment of alternative courses of action, and divide attention among tasks of varying levels of urgency. The DuoWOMBAT provides a good assessment of crew resource management and team coordination for establishing good situational awareness, however, the task is limited to two participants and quantifiable measurement of communication (an important aspect of team performance), is excluded from the data output.

The test was designed to simulate practical military challenges, such as the need for team effectiveness under conditions of operational stress (Breton, Tremblay, and Banbury, 2007). The measure shows good test-retest reliability and predictive validity (Roscoe, Corl, and LaRoche, 2001). Participants are seated side by side at two WOMBAT consoles separated by a partition which increases the communication effort made by the team members. The DuoWOMBAT is composed of a primary target tracking task and three secondary tasks: figure-rotation, quadrant-location, and digit-cancelling. The constructs measured include target tracking, pattern recognition, spatial orientation, and short term memory. Participants are presented with tasks individually and in dual testing phases (Breton, Tremblay, and Banbury). In the target tracking task, participants must maintain two vertical lines on either side of a moving hexagon using their left hand while maintaining a cross centered inside a moving circle with their right hand. The figure-rotation task requires the team members to work together to decide if two figures are identical, mirror images, or different. Each team member controls the rotation of one of the figures. In the quadrant-location task, numbers appear in groups of eight in the four quadrants of the display and participants must cancel the numbers in sequence by pressing the key that corresponds to the quadrant in which the number lies. Finally, in the digit-cancelling task, single digits are displayed sequentially inside a square. Once the third digit is displayed, participants must begin “cancelling” the digits, starting with the first digit displayed, by pressing the digits on the keyboard in the order that they were shown (Breton, Tremblay, and Banbury).

Both tests, the WOMBAT and DuoWOMBAT, were developed by Drs. Jean LaRoche and Stanley Roscoe at Aero Innovation, Inc. Price quotes concerning the equipment and software to utilize this system are available on the Aero Innovation, Inc. website of (www.aero.ca). A 2010 price list is shown at www.aero.ca/e_W_prices_CS.html. To run the DuoWOMBAT, two WOMBAT stations are required; however, the second copy of the software is free of charge. USAARL received a price quote of approximately $60,000 to upgrade its complete (c. 2000) system to 2011 standards. Prices are subject to change and an official quotation must be requested.

The DuoWOMBAT provides a good assessment of crew resource management and team coordination for establishing good situational awareness. However, the task is limited to two participants, the interface is out-of-date in terms of technological display, and quantifiable measurement of communication, an important aspect of team performance, is excluded from the data output. It is important for the experimenter to carefully weigh these factors relative to his/her specific research goals.

Note: The WOMBAT and DuoWOMBAT were purchased c. 2000 by USAARL for discretionary use in its research programs. The decision to purchase these tests was not made by the authors of this review, who were not working at USAARL at the time the tests were purchased. Moreover, USAARL researchers are not required to use the DuoWOMBAT. Therefore, the authors perceive no conflict of interest or bias deriving from the fact that USAARL owned this test prior to the execution of this review by the authors.

Summary of DuoWOMBAT

a. Recommended readings: Roscoe (1993); Roscoe (1997); Roscoe, Corl, and LaRoche (2001); Odle-Dusseau, Bradley, and Pilcher (2010).

b. Team task: Low fidelity multitasking, some information processing and some minor psychomotor aspects.

c. Scenario: Shared situation awareness and crew coordination (e.g., aircrew).

d. Factors studied: Attention to multiple information sources; evaluation of alternatives; establishment of priorities; estimation of probable outcomes of actions.

e. Specific measures: Include shared performance on tasks involving target tracking (attention and psychomotor ability); figure rotation (spatial ability); quadrant location (pattern recognition); and/or digit cancellation (working memory).

f. Strengths: Open source; only test directly relevant to shared display/control (rather than strategic command/control); has been widely used at various sites and seems stable and reliable; tasks well specified in regards to accepted cognitive abilities recognized in neuropsychology.

g. Potential limitations: Simplistic interface – tests basic cognitive abilities; not clear if it can capture communication (e.g., content, frequency); only appears configured for two users; does not have "team of teams" capability; not applicable to strategic command/control; relatively high cost for certain configurations (~$60,000).

<u>New tests in development</u>

After the collection of the literature for this review but prior to the submission of this report, the first author identified a few new preliminary team performance reports (Lum, Sims, and Salas, 2011; Wiese, Pavlas, and Fiore, 2011). These team performance experiments did not directly exploit the existing tests in this review, nor were the experiments primarily intended to develop new computerized team performance test batteries. However, one new computerized test of team performance was identified that is based on the C3Fire test (www.c3fire.org) listed in this review; it is called the C3Conflict (Smith, 2011; Granlund, Smith, and Granlund, 2011). Full evaluation of these new reports was not completed prior to the submission of this report, but a preliminary look at C3Conflict suggests that it has many desirable features which would fit the inclusion criteria of this review and that further attention is warranted as this test matures. While this test is not as developed as the “final seven” recommended in this review, it is mentioned here because it is designed for military applications and its very recent vintage ensures that it is being actively used and will work with the latest computer systems.

Conclusions

Of 54 potential team performance tests identified in this review, the seven tests deemed of greatest interest for military research were (in no order of preference): TANDEM, TPAT, TIDE^2, C^3, NeoCITIES, DDD, AEDGE, and DuoWOMBAT. NeoCITIES is the only test not designed specifically for military applications, but it was included because it has many desirable features and it is suitable for paramilitary (e.g., police) situations and for scenarios relevant to national defense, such as simulating a coordinated emergency response to terrorist attacks on civilian centers. The test most similar to the rudimentary aspects of flight control tasks engaged in by military aviation crewmembers is the DuoWOMBAT. The other six tests focused on various aspects of team performance most relevant to command/control situations, such as handling threats and allocating resources. Among these seven tests, the ones which were judged as most likely to be relevant, readily available, widely/recently used, and relatively mature in terms of validation include NeoCITIES, DDD, and DuoWOMBAT. Of these, the DDD (purchased after review) and DuoWOMBAT (purchased before review) are available for purchase and USAARL has chosen to obtain a copy of each. In addition, a new test was identified (called C3Conflict) after the completion of literature gathering for this review. It is mentioned here because it appears to have many desirable features and should be considered further as additional development and validation is completed.

Recommendations

Researchers studying military or paramilitary team performance should consider the information in this report when seeking to identify the tests most appropriate to the specific needs of the scientific effort being planned. Further use, refinement, validation, and comparison of the existing automated group performance measures are encouraged.

This report identified seven existing tests which appeared most applicable to military research. In addition, the original list of more than 50 potential tests of interest is provided, since some of these other tests may have specific features of importance to a given experiment. For example, the DuoWOMBAT is suitable for an experimental scenario emphasizing shared tactical display/control or shared situation awareness (e.g., crew coordination in the cockpit), while the other six tests described are more appropriate for command/control scenarios.

While no single test will be applicable to all situations, there are obvious drawbacks to publications exploring similar themes in team research separately via dozens of different team performance tests. Single-study, experiment-specific research considerations should be balanced against the multi-study benefits of focusing team research more narrowly on a few key tests used across several laboratories. Such benefits include improved ability for comparison of findings across multiples studies. At least within the restricted realm of command/control tasks, it appears possible to narrow the field of tests considerably. Unfortunately, it is not clear whether each of the final six command/control tasks deemed to be of greatest interest during this review (viz., TANDEM, TPAT, TIDE[2], NeoCITIES, DDD, and AEDGE) are available for immediate use or purchase. It appears that relatively few of the tests of team performance make a successful transition as easily-obtained or commercial off-the-shelf applications. This may lead to a situation where new tests continue to be developed rather than existing tests being improved, validated, and compared. Some of the most important practical questions about team performance measurement will not be answered efficiently by the continued introduction of new measures of team performance. Empirical experience with, and refinement of, existing team performance tests is needed. Until further validation and a head-to-head test comparison is done, a researcher's choice of which test to use will tend to be driven less by the quality of the test's scientific properties than by logistical or psychological factors, such as the test's cost, availability, ease of data administration and analysis, perceived "realism," perceived relevance to the mission or agency, novelty, or place of development (e.g., whether the test was "invented here").

References

Annett, J., Cunningham, D., and Mathias-Jones, P. 2000. A method for measuring team skills. Ergonomics. 43: 1076-1094.

Aptima, Inc. 2010. DDD® 4.1 Adaptable simulation for sociotechnical systems. Brochure.

Aptima, Inc. 2005. Distributed Dynamic Decision-making: The DDD Team Simulation. http://www.aptima.com/Projects/Distributed_Dynamic_Decision_making.html.

Armstrong, J. E. 1994. Designing, testing, and using command control, communications, computers and intelligence (C4I) systems: What causes the disconnects and what can be done about them?. Newport, RI: Naval War College.

Bailey, L., and Thompson, R. 2000. The effects of performance feedback on air traffic control team coordination: A simulation study. Washington, DC: Federal Aviation Administration, Office of Aviation Medicine. FAA Report No DOT/FAA/AM-00/25.

Bailey, L. L., Broach, D. M., Thompson, R. C., and Enos, R. J. 1999. Controller Teamwork Evaluation and Assessment Methodology: A scenario calibration study: Final Report. Oklahoma City, OK: Civil Aeromedical Institute.

Baker, D. P., and Salas, E. 1997. Principles for measuring teamwork: A summary and look toward the future. In M. T. Brannick, E. Salas, and C. Prince (Eds.). Assessment and Measurement of Team Performance: Theory, Methods and Applications. (331-356). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Baker, K., Entin, E. E., See, K., Baker, B. S., Downes-Martin, S., and Cecchetti, J. 2004. Dynamic information and shared situation awareness in command teams. In Proceedings of the 2004 Command and Control Research and Technology Symposium. San Diego, CA.

Banbury, S., Helman, S., Spearpoint, J., and Tremblay, S. 2010. Cracking the bullwhip: team collaboration and performance within a simulated supply chain. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54: 1620-1624. San Francisco, CA.

Banner, M. J. 2004. Assessing team performance and process: Existing tools and tasks in “low-fidelity” environments. Science Applications International Corporation.

Barnes, C., Elliott, L., and Entin, E. 2001. Employing internet2 technology to enable collaborative research and distributed training in complex multi-operation settings. Webnet Journal. 3: 24-31.

Barnes, C., Elliott, L. R., Coovert, M. D., and Harville, D. 2004. Effects of fatigue on simulation-based team decision making performance. Ergometrika. 4: 2-12.

Barnes, C., Elliott, L. R., Tessier, P., and Petrov, P. 2002. Distributed command and control research: Networking multiple simulation platforms. Paper presented at the Seventh International Command and Control Research and Technology Symposium, June, at Monterey, CA. http://www.dodccrp.org/events/7th_ICCRTS/Tracks/pdf/139.PDF.

Barnes, C., Petrov, P. V., and Elliott, L. R. 2002. Agent-based simulation and support of command, control, and communication decisionmaking utilizing a distributed research network. http://dodccrp.org/events/2002/7th_ICCRTS/Tracks/pdf/113.pdf.

Barnes, C., Whitmore, J., Elliott, L., and Harville, D. 2004. Assessing complex team performance in a sustained operations environment. Journal of the International Test and Evaluation Association. 25: 39-44.

Blais, A., Thompson, M. M., and Baranski, J. V. 2002. The effect of individual differences in cognitive styles on decision-making accuracy. Toronto, Ontario, Canada: Defense Research and Development Canada. Report No. DRDC-TORONTO-TR-2003-023.

Bowers, C. A., and Jentsch, F. 2001. Use of commercial, off-the-shelf, simulations for team research. In E. Salas, (Ed.). Advances in Human Performance and Cognitive Engineering Research. Volume 1, 293-317. Kidlington, Oxford, United Kingdom: Elseview Science Ltd.

Bowers, C. A., Salas, E., Prince, C., and Brannick, M. 1992. Games teams play: A method for investigating team coordination and performance. Behavior Research Methods, Instruments, & Computers. 24: 503-506.

Bowers, C. A., Urban, J. M. and Morgan Jr., B. B. 1992. The study of crew coordination and performance in hierarchical team decision making. Orlando, FL: University of Central Florida, Team Performance Laboratory. Report No. TR-92-01.

Brannick, M. T., Prince, A., Prince, C., and Salas, E. 1995. The measurement of team process. Human Factors. 37: 641-651.

Brannick, M. T., Salas, E., and Prince, C. 1997. Team Performance Assessment and Measurement: Theory, Methods and Applications. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Breton, R., Tremblay, S., and Banbury, S. 2007. Measurement of individual and team situation awareness: A critical evaluation of the available metrics and tools and their applicability to command and control environments. Toronto, Ontario, Canada: Defense Research and Development Canada. Report No 2003-283.

Brunswik, E. 1940. Thing constancy as measured by correlation coefficients. Psychological Review. 47: 69-78.

Brunswik, E. 1943. Organismic achievement and environmental probability. Psychological Review. 50: 255-272.

Brunswik, E. 1955. Representative design and probabilistic theory in a functional psychology. Psychological Review. 62: 193-217.

Brunswik, E. 1956. Perception and representative design of experiments. University of California Press: Berkeley, CA.

Cannon-Bowers, J. A., and Salas, E. 1998. Making decisions under stress: Implications for individual and team training. American Psychological Association: Washington, DC.

Canty, A. L., and Schwab, A. 2001. Challenges of virtual teams: The complex effects of personality and turnover on trust, collective efficacy, performance, and member retention. Oxford, MS: University of Mississippi. A286793.

Chapman, T. 2000. The effect of management structure and communication architecture on naturalistic decision making performance. Unpublished Honours Thesis, University of Adelaide, Australia.

Chapman, T., Mills, V., Kardos, M., Stothard, C., and Williams, D. 2002. The use of the Janus Wargame Simulation to investigate naturalistic decision-making: A preliminary examination. Edinburgh, Australia: Land Operations Division, Defense Science and Technology Organization. Report no. DSTO-TR-1372.

Cooke, N. J. 2002. Cognitive Task Analysis for Teams [PowerPoint]. mentalmodels.mitre.org /cog_eng/reference.../Cooke_Teams.ppt.

Cooke, N. J., and Shope, S. M. 2005. Synthetic Task Environments for Teams: CERTT's UAV-STE Handbook on Human Factors and Ergonomics Methods. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas & H. Hendrick (Eds.). Handbook of human Factors and Ergonomics Methods. (46-1 to 46-6). CRC Press, LLC: Boca Raton, FL.

Dahlstrom, N., Dekker, A., Van Winsen, R., and Nyce, J. 2009. Fidelity and validity of simulator training. Theoretical Issues in Ergonomics Science. 10: 305-314.

DeShon, R. P., Brown, K. G., and Greenis, J. L. 1996. Does self-regulation require cognitive resources? Evaluation of resource allocation models of goal setting. Journal of Applied Psychology. 81: 595-608.

DeShon, R. P., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., and Wiechmann, D. 2004. A multiple goal, multilevel model of feedback effects on the regulation of individual and team performance. Journal of Applied Psychology. 89: 1035-1056.

Dube, G., Tremblay, S., Banbury, S., and Rosseau, V. 2010. Team performance and adaptability in crisis management: A comparison of cross-functional and functional teams. In Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting, 54: 1610-1614. San Francisco, CA.

Dwyer, D. J., Hall, J. K., Volpe, C. E., Cannon-Bowers, J. A., and Salas, E. 1992. A performance assessment task for examining tactical decision making under stress. Orlando, FL: Naval Training Systems Center. Special Report No. 92-002.

Elliot, L. R., Miller, J. C., Barnes, C., Dalrymple, M., Brown, L., Whitmore, J., Fischer, J., and Cardenas, R. 2002. Investigation of complex C3 decision making under sustained operations: Issues and analyses. In Proceedings of the 7th International Command and Control Research and Technology Symposium. Monterey, CA.

El-Nasr, S., Jones, R., and McNeese, M. 2004. A scalable and extensible interactive scenario architecture for Distributed Command and Control simulations. In Proceedings of the 2004 Command and Control Research and Technology Symposium. San Diego, CA.

Englund, C. E., Ryman, D. H., Naitoh, P., and Hodgdon, J. A. 1985. Cognitive performance during successive sustained physical work episodes. Behavior Research Methods, Instrumentation & Computers. 17: 75-85.

Entin, E. B., and Entin, E. E. 2000. Assessing Team Situation Awareness in Simulated Military Missions. In Proceedings 44th Annual Meeting of the Human Factors and Ergonomics Society, 44: 73-76. Santa Monica, CA.

Entin, E. E., and Entin, E. B. 2001. Measures for evaluation of team processes and performance in experiments and exercises. In Proceedings of the 6th International Command and Control Research and Technology Symposium, 1-14. Annapolis, MD.

Environmental Tectonics Corporation. 2012. Simulation Training Systems. http://trainingfordisastermanagement.com.

Essens, P., Vogelaar, A., Mylle, J., Blendell, C., Paris, C., Halpin, S., and Baranski J. 2005. Military Command Team Effectiveness: Model and Instrument for Assessment and Improvement. Neuilly-Sur-Seine Cedex, France: Research and Technology Organization. RTO Technical Report No. TR-HRM-087.

Estrada, A., Lawson, B., Kelley, A., and Athy, J. 2010. A Review of Team Performance Measures and Recent Experimental Challenges. Presented at the TTCP Defence Human Systems Symposium, 17-19 May, Sydney, Australia.

Fowlkes, J. E., Dwyer, D. J., Milham, L. M., Burns, J. J., and Pierce, L. G. 1999. Team Skills Assessment: A test and evaluation component for emerging weapon systems. In Proceedings of the 1999 Interservice/Industry Training, Simulation, & Education Conference, 994-1004. Arlington, VA.

Fowlkes, J. E., Lane, N. E., Salas, E., Franz, T., and Oser, R. 1994. Improving the measurement of team performance: The TARGETs methodology. Military Psychology. 6: 47-61.

Fowlkes, J. E., Lane, N. E., Salas, E., Oser, R. L., and Prince, C. 1992. Targets for aircrew coordination training. In Proceedings of the 14th Interservice Industry Training Systems Conference, 342-352. Washington, DC.

Funke, G. J., and Galster, S. M. 2009. The effects of cognitive processing load and collaboration technology on team performance in a simulated command and control environment. International Journal of Industrial Ergonomics. 39: 541-547.

Galster, S. M., Nelson, W. T., and Bolia, R. S. 2005. The utilization of synthetic task environments in C2 domains. Wright Patterson AFB, OH: Air Force Research Laboratory.

Gawron, V. J. 2000. Human Performance Measures Handbook. Mahwah, NJ. Lawrence Erlbaum Associates.

Gaylord, T. L., and Sowell, G. L. 1992. A model-based methodology for evaluating the unit-level air combat mission planning (ACMP) process. Thesis, Air Force Institute of Technology, Wright Paterson Air Force Base, OH.

Geyer-Schulz, A. 1996. The MIT Beer Distribution Game revisited: Machine learning and managerial behavior in a dynamic decision making experiment. Genetic Algorithms and Soft Computing. 8: 658-682.

Go, E., Bos, J. C., and Lamoureux, T. M. 2006. Team Modeling: Survey of Experimental Platforms. Draft Report submitted to Defense Research and Development Canada, Toronto, Ontario, Canada.

Goodwin, J. S., and Franklin, S. G. 1994. The Beer Distribution Game: Using simulation to teach systems thinking. Journal of Management Development. 13: 7-15.

Granlund, R. 2003. Monitoring experiences from command and control research with the C3Fire microworld. Cognition, Technology and Work. 5: 183-190.

Granlund, R., Smith, K., and Granlund, H. 2011. C3Conflict a Simulation Environment for Studying Teamwork in Command and Control. Santa Anna IT Research Institute. http://www.dodccrp.org/events/16th_iccrts_2011/presentations/128.pdf.

Guznov, S., Matthews, G., Funke, G., and Dukes, A. 2011. Use of the Roboflag synthetic task environment to investigate workload and stress responses in UAV operations. Behavior Research Methods. 43: 771-780.

Halpin, S. A. 1996. The Army Command and Control Evaluation System (ACCESS). Fort Leavenworth, MO: U.S. Army Research Institute for the Behavioral and Social Sciences. Report No. 96-04.

Hayes, R. E., Layton, R. L., and Ross, W. A. 1993. The Army Command and Control Evaluation System Documentation. Vienna, VA: Evidence Based Research, Inc.

Hintz, M. 2011. Team Resource Management. http://www.eurocontrol.int/humanfactors/public/standard_page/TRM.html.

Hocevar, S.P., Kemple, W., Kleinman, D., and Porter, G. 1999. Assessments of simulated performance of alternative architectures for command and control: The role of coordination and adaptability. In Proceedings for the 1999 Command & Control Research and Technology Symposium, 123-143. Newport, RI.

Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., and Philips, J. 1995. The multi-level theory of team decision-making: Decision performance in teams incorporating distributed expertise. Journal of Applied Psychology. 80: 292-316.

Hollenbeck, J. R., Sego, D. J., Ilgen, D. R., and Major, D. A. 1991. Team interactive decision exercise for teams incorporating distributed expertise (TIDE2): A program and paradigm for team research. East Lansing, MI: Michigan State University. Report No. 91-1.

Hollenbeck, J. R., Sego, D. J., Ilgen, D. R., Major, D. A., Hedlund, J., and Philips, J. 1997. Team decision-making accuracy under difficult conditions: Construct validation of potential manipulations and measures using TIDE2 simulation. In M. Brannick, E. Salas, and C. Prince (Eds.). New Directions in Team Measurement. (111-136). Jossey Bass: San Francisco, CA.

Holzworth, R. J. 2002. Meta-analysis of team performance accuracy and shared situational awareness in SCUDHunt experiments. Technical Report for ThoughtLink, Inc., Vienna, VA and the Joint C4ISR Decision Support Center, Aberdeen Proving Ground, MD.

Jentsch, F., and Bowers, C. A. 1998. Evidence for the validity of PC-based simulations in studying aircrew coordination. International Journal of Aviation Psychology. 8: 243-260.

Johnston, J. H., Cannon-Bowers, J. A., and Salas, E. 1998. Tactical Decision Making Under Stress (TADMUS): Mapping a Program of Research to a Real World Incident- The USS Vincennes. Edinburgh, United Kingdom: Collaborative Crew Performance in Complex Operational Systems. Report No. TRO-MP-4.

Johnston, J. H., Poirier, J., and Smith-Jentsch, K. A. 1998. Decision making under stress: Creating a research methodology. In Cannon-Bowers J. A. & Salas E. (Eds.). Making Decisions Under Stress: Implications for Individuals and Teams. (39-59). American Psychological Association: Washington, DC.

Jonas, D. (Ed.). 2008. Distributed Team Decision Making. http://www.nasa.gov/centers/ames/research/technology-onepagers/distributed-team-decision.html.

Jones, M. D., McNeese, E. S., Connors, J., Jefferson, T., and Hall, D. 2004. A distributed cognition simulation involving Homeland Security and Defense: The development of NeoCITIES. Paper presented at the 48th annual meeting of the Human Factors and Ergonomics Society, at New Orleans, LA.

Katz, L. C., Kambe, G., Kline, K. F., and Grubb, G. N. 2006. Nonverbal communication and aircrew coordination in army aviation: Annotated bibliography. Arlington, VA: Army Research Institute for the Behavioral and Social Sciences. Report No. 1181.

Knight, D., Durham, C. C., and Locke, E. A. 2001. The relationship of team goals, incentives, and efficacy to strategic risk, tactical implementation, and performance. The Academy of Management Journal. 44: 326-338.

Lamoureux, T. M., Bandali, F., Bruyns Martin, L. M., and Li, Z. 2006. Team modeling: Review of experimental scenarios and computational models. Toronto, Ontario, Canada: Defense Research and Development Canada. Report No. CR2006-092.

Lawson, B., Kelley, A. M., and Athy, J. R. 2011. Team performance measures for use in military research. Slide presentation given at the 82nd Annual Scientific Meeting of the Aerospace Medical Association's, May, in Anchorage, AK.

Lenox, T., Hahn, S., Lewis, M., and Roth, E. 1999. Improving performance: Should we support individuals or teams?. In Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting, 223-227. Santa Monica, CA.

Levchuk, Y. N., Luo J., Levchuk, G. M., Pattipati K. R., and Kleinman, D. L. 1999. A multi-functional software environment for modeling complex missions and devising adaptive organizations. http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA440349.

Lum, H. C., Sims, V. K., and Salas, E. 2011. Low-level predictors of team performance and success. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55: 1457-1461. Las Vegas, NV.

MacMillan, J., Entin, E. B., Hess, K. P., and Paley, M. J. 2004. Measuring performance in a scaled world: Lessons learned from the Distributed Dynamic Decision making (DDD) Synthetic Team Task. In S. G. Schiflett, L. R. Elliott, E. Salas, & M. D. Coovert (Eds.). Scaled worlds: Development, validation, and applications. Brookfield, VT: Ashgate Publishing Co.

Marks, M. A., Sabella, M. J., Burke, C. S., and Zaccaro, S. J. 2002. The impact of cross-training on team effectiveness. Journal of Applied Psychology. 87: 3-13.

McCluskey, M. R., Fowlkes, J. E., Pierce, L. G., and Dwyer, D. J. 1998. Measurement of command/control staff performance in tactical training environments. In Proceedings of the 20th Annual Interservice/Industry Training, Simulation, & Education Conference, 826-833. Orlando, FL.

McNeese, M. D., Bains, P., Brewer, I., Brown, C., Connors, E. S., Jefferson, J. R., Jones, T., and Terell, I. 2005. The NeoCITIES Simulation: Understanding the design and experimental methodology used to develop a team emergency management simulation. In Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting, 591-594. Orlando, FL.

Miller, S. 2001. How Do People Manage Interruptions in Complex Decision Making Tasks? Doctoral Consortium. Presented at Conference on human Factors in Computing Systems. 31 March-5 April, at Seattle, WA.

National Space Biomedical Research Institute. 2010. http://www.nsbri.org/.

Ntuen, C. A. 2006. An experiment in collaborative team decision making in complex information processing environments. Greensboro, NC: North Carolina A&T State University. Project No. 4-41243.

Odle-Dusseau, H. N., Bradley, J. L., and Pilcher, J. J. 2010. Subjective perceptions of the effects of sustained performance under sleep-deprivation conditions. Chronobiology International. 27: 318-333.

Persson, M. and Worm, A. 2002. Information Experimentation in Command and Control. In Proceedings of 2002 Command and Control Research and Technology Symposium. Monterey, CA.

Perla, P. P; Markowitz, M., Nofi, A., Weuve, C., Loughran, J., and Stahl, M. 2000. Gaming and shared situation awareness. Alexandria, VA: Center for Naval Analyses. Report No. CRM D0002722.A2.

Rectenwald, M., Singh, R., Giampapa, J., and Sycara, K. 2003. User Guide for MORSEStation Ranger Operations Simulation. http://www.cs.cmu.edu/~softagents/papers/rectenwald _michael_ 2003.1.pdf.

Reeves, D. L. and Thorne, D. R. 1986. A synopsis of UTC-PAB (Unified Tri-Service Cognitive Performance Assessment Battery). Naval Medical Research Institute Bethesda, MD. http://www.stormingmedia.us/23/2381/A238102.html.

Roscoe, S. N. 1993. An aid in the selection process – WOMBAT. Civil Aviation Training. 4: 48-51.

Roscoe, S. N. 1997. Predicting and enhancing flightdeck performance. In R. Telfer and P. Moore (Eds.). Aviation Training: Learners, Instruction, and Organization. (195-208). Aldershot, United Kingdom: Avebury.

Roscoe, S. N., Corl, L., and LaRoche, J. 2001. Predicting human performance (5th Ed). Saint-Laurent, Canada: Helio Press.

Salas, E., Burke, C. S., Bowers, C. A., and Wilson, K. A. 2001. Team training in the skies: Does crew resource management (CRM) training work? Human Factors. 43: 641-674.

Schoelles, M., and Gray, W. D. 1998. Argus, a System for Varying Cognitive Workload. In Proceedings of the Human Factors an Ergonomics Society 42nd Meeting. Chicago, IL. http://cogworks.cogsci.rpi.edu.

Schraagen, J. M., and Rasker, P. 2003. Team Design. In E. Hollnagel (Ed.). Handbook of Cognitive Task Design. (753-788). Mahwah, NJ: Lawrence Erlbaum Associates.

Shanahan, C., Best, C., Finch, M., and Sutton, C. 2007. Measurement of the behavioural, cognitive, and motivational factors underlying team performance. Fishermans Bend, Victoria, Australia: Defense Science and Technology Organisation. Report No. DSTO-RR-0328.

Shebilske, W. L., Jordan, J. A., Goettl, B. P., and Day, E. 1999. Cognitive and social influences in a training-team protocol for complex skills. Journal of Experimental Psychology: Applied. 5: 227-249.

Simpson, H., and Oser, R. L. 2003. Evaluating large-scale training simulations. Military Psychology. 15: 25-40.

Smith, C. K. 2011. Nett Warrior C3Conflict Experiment: Measuring the effect of battlefield awareness in small units. Monterey, CA: Naval Postgraduate School. Report No. NPS-OR-11-001.

Smith-Jentsch, K. A., Zeisig, R. L., Acton, B., and McPherson, J. A. 1998. Team dimensional training: A strategy for guided self-correction. In J. Cannon-Bowers and E. Salas (Eds.). Making Decisions Under Stress: Implications for individual and team training. (271-298). Washington, DC: APA Press.

Swezey, R. W., Hutcheson, T. D., and Swezey, L. L. 2000. Development of a second-generation computer-based team performance assessment technology. International Journal of Cognitive Ergonomic. 4: 163-170.

Swezey, R. W., Streufert, S., Satish, U., and Siem, F. M. 1998. Preliminary development of a computer-based team performance assessment simulation. International Journal of Cognitive Ergonomics. 2: 163-179.

Sycara, K., Scerri, P., Srinivas, S., and Lewis, M. 2005. Task Allocation in Teams for Launch and Range Operations. In Proceedings of the 11th International Conference on Human Computer Interaction. Las Vegas, NV.

Tactical simulation system (tacsim). 2011. GlobalSecurity.org. http://www.globalsecurity.org/intell/systems/tacsim.htm.

Urban, J. M., Bowers, C. A., Monday, S. D., and Morgan, B. B., Jr. 1995. Workload, team structure, and communication in effective team performance. Military Psychology. 7: 123-135.

Van Berlo, M. P. W. 2004. Improving the quality of team task analysis: Experimental validation of guidelines http://www.iwm-krmc.de/workshops/sim2004/pdf_files/VanBerlo.pdf.

Van Cott, H. P. 2004. Robert W. Swezey (1943-2002). American Psychologist. 59: 46.

Villaneva, R. 2007. Boeing Dedicates Virtual Warfare Center in St. Louis. http://www.boeing.com/news/releases/2005/q2/nr_050509t.html.

Weaver, J., Bowers, C., Salas, E., and Cannon-Bowers, J. 1995. Networked simulations: New paradigms for team performance research. Behavior Research Methods. 27: 12-24.

Weil, S., Hussain, T., Brunyé, T., Sidman, J., and Spahr, L. 2005. The use of massive multi-player gaming technology for military training: A preliminary evaluation. http://openmap.bbn.com/~thussain/??publications/2005_HFES_paper.pdf.

Weise, C. W., Pavlas, D., and Fiore, S. M. 2011. Low-level predictors of team performance and success. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55: 1437-1441. Las Vegas, NV.

Wellens, A. R., and Ergener, D. 1988. The C.I.T.I.E.S. game: A computer-based situation assessment task for studying distributed decision-making. Simulation & Games. 19: 304-327.

Woldring, M., and Issac, A. 1999. Team resource management test and evaluation. European Organisation for the Safety of Air Navigation. http://www.eurocontrol.int/humanfactors/gallery/content/public/docs/DELIVERABLES/HF15%20%28HUM.ET1.ST10.2000-REP-01%29%20Released.pdf.

Wollenbecker, J. M., Kemple, W. G, Kleinman, D. L., Porter, G. R., and Hocevar, S. P. 1999. Using MTWS for Human-in-the-Loop C2 Organizational Experiments. Proceedings for the 1999 Command & Control Research and Technology Symposium. Newport, RI. 144-173.

Department of the Army
U.S. Army Aeromedical Research Laboratory
Fort Rucker, Alabama, 36362-0577
www.usaarl.army.mil

U.S. Army Medical Research and Materiel Command